



**You have downloaded a document from
RE-BUS
repository of the University of Silesia in Katowice**

Title: Desambiguisation des sens des mots et representation lexicale du monde

Author: Wiesław Banyś

Citation style: Banyś Wiesław. (2005). Desambiguisation des sens des mots et representation lexicale du monde. "Neophilologica" (T. 17, (2005), s. 57-76).



Uznanie autorstwa - Użycie niekomercyjne - Bez utworów zależnych Polska - Licencja ta zezwala na rozpowszechnianie, przedstawianie i wykonywanie utworu jedynie w celach niekomercyjnych oraz pod warunkiem zachowania go w oryginalnej postaci (nie tworzenia utworów zależnych).



UNIwersYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego

Wiesław Banyś
Université de Silésie
Katowice

Désambiguïsation des sens des mots et représentation lexicale du monde

Abstract

Wiesław Banyś discussed the natural polysemy of the lexicon of natural languages, perceived as one of the major problems of the automatic translation. He also presented a solution for the problem of disambiguation of ambiguous expressions – it pivots on the object-oriented description, enriched with a transfer of various possible translations in a target language to the possible ambiguity of a source language. Finally, the author showed differences in the representations of the world in Polish and French, on the example of some Polish renderings of the French verb *gagner*.

Keywords

Polysemy, ambiguity, disambiguation, object-oriented description, object classes, automatic translation

Nous allons nous concentrer dans ce qui suit sur l'un des problèmes majeurs de la traduction assistée par ordinateur (dorénavant : TAO) qui est constitué par la polysémie des mots en langue naturelle et par conséquent la nécessité de la désambiguïsation des sens des mots (dorénavant : WSD) automatisable et efficace au cours de l'interprétation du texte de départ pour que le système puisse, entre autres, fournir la bonne génération de leurs équivalents dans la langue d'arrivée.

Cela nous permettra de voir aussi, à partir de la description des sens du verbe FR *gagner*, les différentes représentations du monde que donne le verbe français en question et ses équivalents en polonais.

Nous allons tout d'abord donner quelques exemples de la réelle traduction automatique fournie par certains systèmes de traduction automatique : *Quelques exemples de difficultés de départ* (Sec. 1).

Cela nous permettra de discuter :

- d'une part, la question des distinctions de sens en général, ensuite la discussion de celles qu'on trouve dans les dictionnaires papier traditionnels et celles qui sont à établir à l'aide des procédures descriptives proposées,
- et, d'autre part, les conséquences qui en résultent pour la confection d'un dictionnaire bilingue servant à la traduction assistée par ordinateur (TAO) du FR vers le PL / du PL vers le FR en particulier.

Ensuite, nous présenterons l'éventail des sens du verbe FR *gagner* tel qu'on trouve dans le *Trésor de la langue française* et *Le Grand Robert de la Langue Française* : «Gagner» à gagner? (Sec. 2).

La description satisfaisant les exigences d'une présentation contrastive opératoire des sens, faite sur l'exemple du verbe *gagner* et de ses équivalents polonais, terminera notre exposé : *Gagnons* «gagner» (Sec. 3).

1. Quelques exemples de difficultés de départ

Quand on regarde les traductions fournies par la majorité, pour ne pas dire tous, les systèmes de traduction appelée automatique ou assistée par ordinateur (TAO), on voit tout de suite (cf. à cet égard p.ex. déjà les travaux de Y. Bar-Hillel (1960) sur la TAO) que l'un des problèmes majeurs que ces systèmes doivent affronter est la polysémie des mots qui freine trop souvent des traductions adéquates des textes (cf. à ce propos une littérature abondante traitant de ce problème, p.ex. N. Ide, J. Véronis, 1998).

Voilà des exemples de traduction proposée par certains systèmes actuels où la polysémie montre quelques facettes :

He always wears a jacket and tie to work ⇒ ^{*}*Il porte toujours une veste et un lien pour travailler*

He offered to walk her home ⇒ ^{*}*Il a offert de marcher sa maison*

A heavily loaded car really came in ⇒ ^{*}*Cieężko załadowany samochód rzeczywiście przychodził* [*Une voiture fortement chargée est vraiment venue*]

The car crashed into the building ⇒ ^{*}*Samochód huknął do budującego* [*La voiture s'est écrasée dans le bâtiment*]

La vitre s'est cassée ⇒ ^{*}*Szyba sie złamała* [*The pane broke*]

Gagner d'une longueur ⇒ ^{*}*to gain a length* [*⇒ to win by a length*]

La WSD est nécessaire non seulement dans les tâches de traduction assistée par ordinateur (TAO), mais aussi dans d'autres tâches liées à un traitement

computationnel de la langue largement conçu, p.ex. la recherche de l'information et la navigation hypertextuelle sur Internet, l'analyse automatique des mots en parties du discours, le résumé automatique, le traitement automatique de la parole.

Il est naturel que l'on doive, vu la complexité de la tâche, avoir recours aux différents paramètres, « sources d'information », pour essayer de mener à bien la WSD (v. à ce propos entre autres, S. McRoy, 1992; H. Ng, H. Lee, 1996; Y. Wilks et al., 1996; Y. Wilks, M. Stevenson, 1998b).

Les caractéristiques générales qui nous guident donc dans la WSD sont de différent type, p.ex.: parties du discours et leur genre (p.ex. *un livre* vs *une livre*, *dîner* vs *le dîner*), synonymes, antonymes, schémas syntaxico-sémantiques (p.ex. *jouer avec* vs *jouer de*), structures prédicats – arguments différentes (p.ex. *lire un livre* vs *prendre un livre*) dont les restrictions de sélection sémantique sur les arguments des prédicats (*événement* vs *objet*, p.ex.), mais aussi les restrictions de sélection lexicale sur les arguments des prédicats (p.ex. types d'abstrais: *observer un match* (PL: *oglądać mecz*) vs *observer les règles* (PL: *przestrzegać reguł*), ce qui est le résultat de, ou si l'on veut, entraîne, l'appartenance aux classes sémantiques et les règles d'héritage sémantique différentes et est, par conséquent, une manifestation de relations sémantiques syntagmatiques et paradigmatisques différentes dont le cas extrême, le tout étant un continuum, est représenté par les « collocations » ou expressions plus ou moins figées (cf. p.ex. à ce propos le livre très éclairant et précisant beaucoup de points généralement confondus concernant les expressions figées de G. Gross (1996)).

Nous y ajouterons encore, comme facteur important, à prendre en considération avec les précédents, du point de vue de l'opérationnalité des dictionnaires électroniques utilisés dans la TAO bilingue, les équivalents dans l'autre langue: langue-cible (cf. à ce propos p.ex. les travaux de I. Dagan et A. Itai, 1994).

Autant dire que, du point de vue présenté, il y a autant de sens différents d'un mot dans la langue source que de ses traductions différentes dans la langue d'arrivée et d'ensembles de marqueurs syntagmatiques différents qui les distinguent les uns des autres dans la langue de départ.

Dans une approche du type « efficacité avant tout », l'importance du second paramètre peut être réduite d'une manière significative dans la mesure où un mot peut avoir différents ensembles de marqueurs syntagmatiques dans la langue de départ, mais être traduit néanmoins de la même manière dans la langue d'arrivée, ce qui est d'ailleurs un reflet de miroir dialectique de l'approche appliquée changeant d'optique en cours de route (cela suppose, du point de vue de chacune des langues, une communauté sémantique opérationnelle des constructions à un niveau plus abstrait d'analyse, type de schème invariant (cf. J.-P. Desclés, W. Banyś, 1997) ou règle heuristique (cf.

W. Banyś, 2002a, b) ou encore de cour sémantique du mot, menant en fait au traitement monosémique des mots à la façon d'A. Wierzbicka (1989) qui sont à relever par la suite, p.ex. FR: *Gagner de l'argent à un jeu / gagner un procès* vs PL: *wygrywać*.

Ce type d'approche ressemble aussi d'une certaine manière, sans y être identique, à la position en la matière d'A. Kilgarriff (1993) dont son fameux *I don't believe in word senses* (A. Kilgarriff, 1997) est le couronnement, qui niait la possibilité objective de l'établissement d'un inventaire défini de sens d'un mot dans une langue donnée, vu, entre autres, le fait que les lexicographes sont, d'après A. Kilgarriff, obligés de décrire les mots comme s'ils possédaient tous des ensembles discrets, sans intersections, de sens. Cela entraîne comme conséquence le fait qu'il est souvent arbitraire d'assigner le sens du mot dans un emploi concret dans le discours à l'un des sens du mot présentés dans le dictionnaire.

C'est pourquoi d'ailleurs, d'après A. Kilgarriff, p.ex. P. Hanks (1994), qui analysait les sens de *climb* en anglais ou Ch. Fillmore et B.T.S. Atkins (1992), analysant le champ sémantique de *risk* en anglais prônaient le recours à l'analyse prototypique ou l'analyse en termes de *frames* comme cadres plus riches à la description des sens des mots; on sait qu'il y avait pour ce type d'approches des raisons plus fondamentales: changement de paradigme de voir l'essence de la langue en tout: du logique au cognitif, du partiel à l'holistique.

L'implication de la position d'A. Kilgarriff (1993) est que les sens de mots ne sont définis que par rapport à un ensemble d'intérêts du lexicographe, c'est d'ailleurs, sous une forme beaucoup plus pondérée, ce que disait p.ex. entre autres Y. Wilks (1972), et que, d'autre part, l'ensemble de sens d'un mot défini dans un dictionnaire peut ou ne peut pas correspondre à l'ensemble de sens qui est relevant pour une application du traitement computationnel de la langue (v. à cet égard aussi une discussion critique éclairante des thèses radicales d'A. Kilgarriff (1993, 1997) dans Y. Wilks (1997)).

Aussi bien l'un que l'autre et beaucoup d'autres tombent d'accord que l'assignement d'un sens donné à une occurrence d'un mot dans le texte est relative au choix de sens qu'un dictionnaire présente, à ceci près qu'A. Kilgarriff soutient qu'il est très souvent impossible de faire un tel assignement, en glissant sans doute ainsi de la question quels sont et combien ils sont les sens contenus dans un dictionnaire vers la question de nouveaux sens ou des sens ponctuellement créés dans le discours par des procédés rhétoriques du type métaphore, métonymie, etc., et Y. Wilks distingue, par contre, nettement les deux problèmes, au statut tout à fait distinct, mais la question est en fait de savoir comment on établit les différents sens, et leur nombre, dans un dictionnaire.

Et l'on peut imaginer que la réponse sera différente suivant les tâches à accomplir, dont l'une des facettes est aussi le fait de savoir si l'on parle d'un dictionnaire monolingue, bilingue, papier, électronique, etc.

Notre position ressemble à l'approche relativiste d'A. Kilgarriff (1993, 1997) dans la mesure où elle insiste sur les analyses de corpus pour relever les emplois réels d'un mot et où elle relativise le nombre de sens d'un mot donnée de la langue-source dans un dictionnaire électronique à un « intérêt » particulier de la description, qui est une traduction dans une autre langue, ce qui se réduit au nombre de traductions différentes que le mot en question reçoit dans la langue-cible, mais en diffère, parce qu'elle ajoute encore aux éléments de décision le paramètre d'ensembles de caractéristiques syntagmatiques. Cela permet d'ailleurs de traiter ce type d'approche comme une sorte d'aide à l'analyse du bien-fondé des distinctions de sens faites dans un dictionnaire traditionnel de la langue-source.

Cela veut dire que notre approche se réduit à quelques démarches de décision :

- on établit la concordance des emplois du mot étudié dans un large corpus,
- on répartit les concordances trouvées en ensembles dont les éléments ont plus de traits syntaxiques, sémantiques et lexicaux en commun qu'avec les éléments appartenant à d'autres ensembles,
- on applique l'approche orientée objets (cf. p.ex. G. Gross, 1997, 1999a, b ; W. Banyś, 1997, 2002a, b) pour analyser et regrouper les caractéristiques communes,
- on y assigne les traductions en langue-cible,
- on réorganise, si nécessaire, les ensembles établis en fonction des résultats de la traduction,
- on code les conclusions de l'analyse dans l'un des formats descriptifs (se situant dans l'approche orientée objets, on applique ici l'une des variantes de ses fiches descriptives).

2. « Gagner » à gagner ?

C'est cette approche donc qu'on va appliquer à la description des sens du verbe FR *gagner* – le corpus a été limité, pour les besoins de la démonstration, au texte du *Trésor de la langue française* et *Le Grand Robert de la Langue Française*.

Cela nous a permis d'effectuer l'inventaire des types suivants de sens relevés dans les textes cités :

1. *Gagner de l'argent, une fortune, un salaire*
2. *Je n'y ai gagné que des coups ; J'ai gagné à la bataille du 29 janvier, livrée entre l'autorité législative et la prérogative présidentielle, trois ans de prison*

3. *Je me suis promené sans manteau, j'y ai gagné un bon rhume ; L'hiver a été le plus rude qu'on ait vu depuis quarante ans. J'y ai gagné la sciatique*
4. *Gagner de l'argent à un jeu, aux courses ; Gagner les consommations aux cartes*
5. *Gagner un procès, sa cause, une bataille, la guerre, le match, un pari, la partie, une partie d'échecs, un tournoi de bridge*
6. *Son parti avait gagné dix sièges, les suffrages, les voix aux élections ; Gagner le prix pendant le concours ; Cet athlète a gagné sa sélection*
7. *Je te gagnerai bien un jour ! Il m'a gagné sur cent mètres*
8. *Jospin aurait pu gagner Chirac ; Le mal gagne parfois avec le bien, les ordinateurs gagnent avec les machines à écrire ; Il peut gagner avec sa maladie*
9. *L'enfant qui gagne x... centimètres/kilos ; Gagner de l'appétit, de belles manières, de l'influence ; Gagner le bonheur éternel, l'éternité ; Gagner de la gloire, de la notoriété*
10. *Gagner la faveur, l'amitié, la confiance, l'estime, la bienveillance, les bonnes grâces... de quelqu'un ; Il a su gagner la bienveillance du public par sa simplicité, sa gentillesse*
11. *Sa simplicité, sa gentillesse lui a gagné de la gloire, de la notoriété, de l'influence, la faveur, l'amitié, la confiance, l'estime, la bienveillance, les bonnes grâces... de quelqu'un ; Sa simplicité, sa gentillesse lui a gagné la bienveillance du public*
12. *Gagner les coeurs du public ; Gagner ses galons sur le champ de bataille*
13. *Sa simplicité lui a gagné les coeurs du public*
14. *L'important dans ce procès est de gagner du temps*
15. *Cette situation vous fera gagner une heure*
16. *Celle-ci monte, émet son bourgeon terminal pour gagner en hauteur, ses bourgeons axillaires pour gagner en largeur*
17. *Vous avez tout à gagner à ce choix, à faire ce choix ; Il gagnera à s'associer à vous, à cette association ; Vous gagneriez à ce que cette affaire ne s'ébruite pas ; Cette toile gagne à être vue à la lumière du jour ; Ce vin gagnera à vieillir ; Ce garçon gagne à être connu*
18. *Et ils y gagnent d'être mieux connus sans être moins aimés ; Vous y gagnerez d'être enfin tranquille ; Par sa dignité, il a gagné d'être respecté jusque dans ses échecs*
19. *Vous devez gagner les abris en cas d'alarme ; On l'arrêta avant qu'il eût gagné la frontière ; Le navire gagna la haute mer, le large ; Il nous fallut trois heures pour gagner à pied le refuge ; Gagner la porte, gagner la sortie*
20. *Gagner le ciel, le paradis par ses oeuvres, par ses vertus*
21. *Le feu gagnait déjà la maison voisine ; L'eau a gagné le second étage ; La contagion gagna plusieurs quartiers ; L'incendie, l'inondation gagne ; La gangrène a gagné ; L'insurrection, la grève gagne tout le pays*

22. *Le feu gagnait partout ; L'incendie, l'inondation gagne ; La gangrène a gagné ; L'insurrection, la grève gagne partout*
23. *Ces idées gagnent toute la population*
24. *Ces idées gagnent parmi la population*
25. *Gagner des indulgences, l'intercession de la Vierge, son salut*
26. *Gagner quelqu'un par de bons procédés, par la franchise ; Il a su gagner le public par sa simplicité, sa gentillesse ; L'art de gagner les hommes, les esprits ; Gagner de nouveaux partisans, des fidèles, des prosélytes ; Nous l'avons enfin gagné à notre cause ; Gagner quelqu'un par ses caresses, ses cajoleries*
27. *La nuit, l'orage, la marée nous gagne*
28. *Le froid, le sommeil, la faim, la fatigue commençaient à le gagner ; Agitation, impression peur, pitié qui gagnent qqn ; Le froid l'avait déjà gagné ; L'ennui gagna l'auditoire*
29. *Sa tristesse, sa confiance me gagnait ; Un esprit, un ton particulier qui gagne le lecteur*
30. *Je me sentais gagné par son optimisme, par cette gaieté communicative ; Il se sentait gagné par un étrange malaise ; Être gagné par l'émotion, par les larmes*
31. *Gagner à la course, aux échecs*

3. Gagnons «gagner»

Les analyses syntaxiques, sémantiques et lexicales des constructions relevées ont permis de fournir l'approximation suivante des sens de *gagner* vus à travers ses traductions polonaises où, entre les crochets [...], on note les classes d'objets plus générales et entre les < ... > les classes d'objets plus spécifiques, tout étant relatif, impliquées.

Gagner de l'argent, une fortune, un salaire

X – [ANM] – gagner – Y – [<avantage> : <profit financier> : <argent>] [ACTIVITE: <travail>] zarabiac

Je n'y ai gagné que des coups ; J'ai gagné à la bataille du 29 janvier, livrée entre l'autorité législative et la prérogative présidentielle, trois ans de prison

X – [ANM] – gagner – Y – [<non-avantage> : <mauvais>] [ACTIVITE: ALL] zarabiac

Je me suis promené sans manteau, j'y ai gagné un bon rhume ; L'hiver a été le plus rude qu'on ait vu depuis quarante ans. J'y ai gagné la sciatique

X – [ANM] – gagner – Y – [<non-avantage> : <mauvais> : <mala-die>] dostać

Gagner de l'argent à un jeu, aux courses; Gagner les consommations aux cartes

X – [ANM/INSTR du jeu] – gagner – Y – [<avantage> : <profit financier/non-financier> : <enjeu> <argent> <quantification d'argent>] [ACTIVITE: jeu] wygrywać

Gagner un procès, sa cause, une bataille, la guerre, le match, un pari, la partie, une partie d'échecs, un tournoi de bridge

X – [ANM] – gagner – Y – [<affrontement, compétition, rivalité, procès, jeu, etc.>] wygrywać

Son parti avait gagné dix sièges, les suffrages, les voix aux élections; Gagner le prix pendant le concours; Cet athlète a gagné sa sélection

X – [CONC] – gagner – Y – [<enjeu d'affrontement compétition, rivalité, procès, jeu, etc.> <demandant bcp d'efforts pour l'obtenir>] [ACTIVITE: <affrontement, compétition, rivalité, procès, jeu, etc.>] zdobywać

[Enjeu – fonction du type de compétition: partie – tout: *concours: prix, place; élections: siège, votes, voix*]

Je te gagnerai bien un jour! Il m'a gagné sur cent mètres

X – [ANM] – gagner – Y – [ANM] wygrywać z kimś

Jospin aurait pu gagner Chirac; Le mal gagne parfois avec le bien, les ordinateurs gagnent avec les machines à écrire; Il peut gagner avec sa maladie

X – [ALL] – gagner – avec – Y – [ALL] wygrywać z kimś

L'enfant qui gagne x... centimètres/kilos; Gagner de l'appétit, de belles manières, de l'influence; Gagner le bonheur éternel, l'éternité; Gagner de la gloire, de la notoriété

X – [ANM] – gagner – du/de la/des/le/la/les Y – [<avantage> : <profit non-financier> : <qualité de X>] zyskać coś

Gagner la faveur, l'amitié, la confiance, l'estime, la bienveillance, les bonnes grâces... de quelqu'un; Il a su gagner la bienveillance du public par sa simplicité, sa gentillesse

X – [ANM] – gagner – du/de la/des/le/la/les Y – [<avantage> : <profit non-financier> : <qualité de Y>] { par Z <qualité de X> } zyskać coś dzięki czemuś

Sa simplicité, sa gentillesse lui a gagné de la gloire, de la notoriété, de l'influence, la faveur, l'amitié, la confiance, l'estime, la bienveillance, les

bonnes grâces... de quelqu'un ; Sa simplicité, sa gentillesse lui a gagné la bienveillance du public

Z – [qualité de X>] – gagner – du/de la/des/le/la/les Y – [<avantage> : <profit non-financier> : <qualité de Y>] zyskać coś komuś

Gagner les coeurs du public, Gagner ses galons sur le champ de bataille
X – [ANM] – gagner – Z – [:<avantage> : <profit non-financier> <demandant bcp d'efforts pour l'obtenir>] { par Z <qualité de X> } zdobywać

Sa simplicité lui a gagné les coeurs du public

Z – [<qualité de X>] – gagner – Z – [ABSTR :<avantage> : <profit non-financier> <demandant bcp d'efforts pour l'obtenir>] zdobywać

L'important dans ce procès est de gagner du temps

X – [ANM] – gagner – du/de la/des Y – [<temps>] zyskać na

Cette situation vous fera gagner une heure

X – [ANM] – gagner – Y – [<temps>] zyskać [czas]

Celle-ci monte, émet son bourgeon terminal pour gagner en hauteur, ses bourgeons axillaires pour gagner en largeur

X – [ALL] – gagner en – Y – [<avantage> : <profit non-financier> : <qualité de X>] zyskać na

Vous avez tout à gagner à ce choix, à faire ce choix ; Il gagnera à s'associer à vous, à cette association ; Vous gagneriez à ce que cette affaire ne s'ébruite pas ; Cette toile gagne à être vue à la lumière du jour ; Ce vin gagnera à vieillir ; Ce garçon gagne à être connu

X – [ALL] – gagner à – faire/être Y/à ce que Z/ – zyskać na/że/ kiedy/jeśli

Et ils y gagnent d'être mieux connus sans être moins aimés ; Vous y gagnerez d'être enfin tranquille. Par sa dignité, il a gagné d'être respecté jusque dans ses échecs

X – [ANM] – gagner de – INFIN – zyskać, że / na / kiedy / jeśli

Vous devez gagner les abris en cas d'alarme ; On l'arrêta avant qu'il eût gagné la frontière ; Le navire gagna la haute mer, le large ; Il nous fallut trois heures pour gagner à pied le refuge ; Gagner la porte, gagner la sortie

X – [CONC : <voyageur> <véhicule>] – gagner – Y – [<espace>] dotrzeć do

Gagner le ciel, le paradis par ses oeuvres, par ses vertus

X – [ANM] – gagner – Y – [<espace>] par Z <qualité de X> dostać się do

Le feu gagnait déjà la maison voisine ; L'eau a gagné le second étage ; La contagion gagna plusieurs quartiers ; L'incendie, l'inondation gagne ; La gangrène a gagné ; L'insurrection, la grève gagne tout le pays

X – [ABSTR:<incendie, maladie, troubles, chômage>] – gagner
– Y – [<espace>] rozszerzać się na

Le feu gagnait partout ; L'incendie, l'inondation gagne ; La gangrène a gagné ; L'insurrection, la grève gagne partout

X – [ABSTR:<incendie, maladie, troubles, chômage>] – gagner
rozszerzać się

Ces idées gagnent toute la population

X – [ABSTR:<idées> >] – gagner – Y – [<ANM>] szerzyć się wśród

Ces idées gagnent parmi la population

X – [ABSTR:<idées> >] – gagner szerzyć się wśród

Gagner des indulgences, l'intercession de la Vierge, son salut

X – [ANM] – gagner – Y – [ABSTR:<avantage> : <profit non-financier religieux> <demandant des efforts pour l'obtenir>]
uzyskać

Gagner quelqu'un par de bons procédés, par la franchise ; Il a su gagner le public par sa simplicité, sa gentillesse ; L'art de gagner les hommes, les esprits ; Gagner de nouveaux partisans, des fidèles, des prosélytes ; Nous l'avons enfin gagné à notre cause ; Gagner quelqu'un par ses caresses, ses cajoleries

X – [ANM] – gagner – Y – [ANM] [<demandant des efforts pour l'obtenir>] { par Z <qualité de X> } pozyskiwać kogoś dzięki czemuś

La nuit, l'orage, la marée nous gagne

X – [<phen meteo>] – gagner – Y [<espace>] zbliżać się

Le froid, le sommeil, la faim, la fatigue commençaient à le gagner ; Agitation, impression peur, pitié qui gagnent qqn ; Le froid l'avait déjà gagné ; L'ennui gagna l'auditoire

X – [<émotion, rire> de Y> <sentiment/sensation négatifs> : <peur, angoisse, ennui, découragement, fatigue, froid, sommeil>] – gagner
– Y [ANM] ogarniać

Sa tristesse, sa confiance me gagnait ; Un esprit, un ton particulier qui gagne le lecteur

X – [<état, émotion, qualité, rire> de V] – gagner – Y [ANM] udzielać się komu

Je me sentais gagné par son optimisme, par cette gaieté communicative ; Il se sentait gagné par un étrange malaise ; Être gagné par l'émotion, par les larmes

X – [*<ANM>*] – être/se sentir – gagné par Y – [*<état, émotion, qualité, rire>*] *czuć się opanowanym przez*

Gagner à la course, aux échecs

X – [*ANM*] – gagner – à – Y – [*<jeu, compétition, élections, etc.>*] *wygrywać*

Concurrent qui gagne d'une longueur, d'une tête

X – [*ANM*] – gagner – de – Y – [*ABSTR: <indication de distance>*] { pendant [*<jeu, compétition, élections, etc.>*] *wygrywać o*

Les résultats de ces analyses peuvent être représentés sous forme d'une fiche descriptive (voir pp. 68–72) ne reprenant ici que les caractéristiques relevées ci-dessus, sans mentionner les éléments tels que p.ex. synonymes, antonymes ou appartenance aux classes d'héritage sémantique (cf. p.ex. W. B a n y ś, 2002a, b).

Nous présentons ci-dessous quelques commentaires concernant l'analyse effectuée et ses enjeux théoriques et plus particulièrement :

- I – la question d'assignement possible d'un sens concret apparaissant dans le discours à un sens défini dans le dictionnaire,
- II – la question du nombre de sens de mot dans un dictionnaire monolingue et bilingue,
- III – la question de la validité de la répartition des sens des mots de la langue source suivant leur traduction en langue cible,
- IV – la question de la création des classes d'objets.

Quant à la première question, il est tout à fait évident qu'il faut la distinguer de la seconde, sinon, et c'est parfois le cas d'A. Kilgariff (1993), semble-t-il, on ne pourra pas affecter un sens nouveau ou occasionnel à l'un des sens existants dans le dictionnaire, parce que c'est justement un sens nouveau ou occasionnel.

Remarquons à ce propos que Y. Wilks p.ex. dans sa sémantique de préférences a consacré une place privilégiée à l'étude des divergences dans l'interprétation des occurrences par rapport aux sens déterminés, qui sont à considérer comme norme lexicale ou préférence, dans le dictionnaire et s'il y a un sens d'occurrence qui diverge de ceux que l'on trouve dans le dictionnaire soit on essaye de l'assigner à l'un d'eux (cf. p.ex. les procédures proposées dans Y. Wilks, 1972) soit, si ce n'est pas possible, on construit un sens distinct (cf. à ce propos les procédures proposées dans Y. Wilks, 1978).

La deuxième et la troisième question sont intimement liées dans l'approche présentée. Nous clarifierons davantage notre position en analysant l'un des exemples étudiés par Y. Wilks (1972: 12).

NP1 CO	Verbe	Prép.	NP2 CO	Prép.	NP3 CO	NP1 cas	NP1 CO	Verbe	Prép.	NP2 cas	NP2 CO	Prép.	NP3 cas	NP3 CO
[ANM]	<i>gagner</i>	[PART/ ART]	[< avantage > : < profit non- financier > : < qualité de X >]	–	–	[NOM]	[ANM]	<i>zyskać</i>	–	[ACC]	[< avantage > : < profit non- financier > : < qualité de X >]	–	–	–
[ANM]	<i>gagner</i>	[PART/ ART]	[< avantage > : < profit non- financier > : < qualité de Y >]	<i>par</i>	Z < qualité de X >	[NOM]	[ANM]	<i>zyskać</i>	–	[ACC]	[< avantage > : < profit non- financier > : < qualité de Y >]	<i>dzięki</i>	[DAT]	Z < qualité de X >
Z – [ABSTR : < qualité de X >]	<i>gagner</i>	[PART/ ART]	[< avantage > : < profit non- financier > : < qualité de Y >]	<i>à</i>	[ANM]	[NOM]	Z – [ABSTR : < qualité de X >]	<i>zyskać</i>	–	[ACC]	[< avantage > : < profit non- financier > : < qualité de Y >]	–	[DAT]	[ANM]
[ANM]	<i>gagner</i>	[PART]	[< temps >]	[FUNC]	[ACTIVITE : < ALL >]	[NOM]	[ANM]	<i>zyskać</i>	<i>na</i>	[INSTR]	[< temps >]	–	–	–
[ANM]	<i>gagner</i>	[DET]	[< temps >]	[FUNC]	[ACTIVITE : < ALL >]	[NOM]	[ANM]	<i>zyskać</i>	–	[ACC]	[< temps >]	–	–	–
[ALL]	<i>gagner</i>	<i>en</i>	[< avantage > : < profit non- financier > : < qualité de X >]	–	–			<i>zyskać</i>	<i>na</i>	[INSTR]	[< avantage > : < profit non- financier > : < qualité de X >]	–	–	–
[ALL]	<i>gagner</i>	<i>à</i>	[INFIN]	–	–	[NOM]	[ALL]	<i>zyskać</i>	<i>że/kiedy/ jeśli</i>	Σ	–	–	–	–
[ALL]	<i>gagner</i>	<i>à ce que</i>	Σ	–	–	[NOM]	[ALL]	<i>zyskać</i>	<i>na tym, że</i>	Σ	–	–	–	–
[ANM]	<i>gagner</i>	<i>de</i>	[INFIN]	–	–	[NOM]	[ALL]	<i>zyskać</i>	<i>że/kiedy/ jeśli</i>	Σ	–	–	–	–

[CONC]	<i>gagner</i>	[DET]	[<enjeu d'affrontement compétition, rivalité, procès, jeu, etc.> <demandant bcp d'efforts pour l'obtenir>]	[FUNC]	[ACTIVITE: <affrontement, compétition, procès, jeu, etc.>]	[NOM]	[CONC]	<i>zdobywać</i>	-	[ACC]	[<enjeu d'affrontement compétition, rivalité, procès, jeu, etc.> <demandant bcp d'efforts pour l'obtenir>]	[FUNC]	[FUNC]	[ACTIVITE: <affrontement, compétition, rivalité, procès, jeu, etc.>]
[ANM]	<i>gagner</i>	[DET]	[<avantage> : <profit non-financier> <demandant bcp d'efforts pour l'obtenir>]	<i>par</i>	Z <qualité de X>	[NOM]	[ANM]	<i>zdobywać</i>	-	[ACC]	[<avantage> : <profit non-financier> <demandant bcp d'efforts pour l'obtenir>]	<i>dzięki</i>	[DAT]	Z <qualité de X>
Z - [<qualité de X>]	<i>gagner</i>	[DET]	[<avantage> : <profit non-financier> <demandant bcp d'efforts pour l'obtenir>]	<i>à</i>	[ANM]	[NOM]	Z - [<qualité de X>]	<i>zdobywać</i>	-	[ACC]	[<avantage> : <profit non-financier> <demandant bcp d'efforts pour l'obtenir>]	-	[DAT]	[ANM]
[<phénométéo>]	<i>gagner</i>	[DET]	[<espace>]	-	-	[NOM]	[<phénométéo>]	<i>zblizać się</i>	<i>do</i>	[GEN]	[<espace>]	-	-	-
[ANM]	<i>gagner</i>	[DET]	[<avantage> : <profit financier> : <argent>]	[FUNC]	[ACTIVITE: <travail>]	[NOM]	[ANM]	<i>zarabiać</i>	-	[ACC]	[<avantage> : <profit financier> : <argent>]	[FUNC]	[FUNC]	[ACTIVITE: <travail>]
[ANM]	<i>gagner</i>	[DET]	[<non-avantage> : <mauvais>]	[FUNC]	[ACTIVITE: <ALL>]	[NOM]	[ANM]	<i>zarabiać</i>	-	[ACC]	[<non-avantage> : <mauvais>]	[FUNC]	[FUNC]	[ACTIVITE: <ALL>]

NP1 CO	Verbe	Prép.	NP2 CO	Prép.	NP3 CO	NP1 cas	NP1 CO	Verbe	Prép.	NP2 cas	NP2 CO	Prép.	NP3 cas	NP3 CO
[ANM/ INSTR du jeu]	<i>gagner</i>	[DET]	[< avantage > : < profit financier/non- financier > : < enjeu > < argent > < quantité d'argent >]	[FUNC]	[ACTIVITE: < jeu >]	[NOM]	[ANM/ INSTR du jeu]	<i>wygrywać</i>	-	[ACC]	[< avantage > : < profit financier/non- financier > : < enjeu > < argent > < quantité d'argent >]	[FUNC]	[FUNC]	[ACTIVI- TE: < jeu >]
[ANM]	<i>gagner</i>	[DET]	[< affronte- ment, compétition, rivalité, procès, jeu, etc. >]	-	-	[NOM]	[ANM]	<i>wygrywać</i>	-	[ACC]	[< affronte- ment, compétition, rivalité, procès, jeu, etc. >]	-	-	-
[ANM]	<i>gagner</i>	[DET]	[ANM]	[FUNC]	[ACTIVITE: < jeu >]	[NOM]	[ANM]	<i>wygrywać</i>	z	[ABL]	[ANM]	[FUNC]	[FUNC]	[ACTIVI- TE: < jeu >]
[ALL]	<i>gagner</i>	avec	[ALL]	-	-	[NOM]	[ANM]	<i>wygrywać</i>	z	[ABL]	[ALL]	-	-	-
[ANM]	<i>gagner</i>	[FUNC]	[< jeu, compétition, élections, etc. >]	-	-	[NOM]	[ANM]	<i>wygrywać</i>	-	[ACC]	[< jeu, compétition, élections, etc. >]	-	-	-
[ANM]	<i>gagner</i>	de	[< indication de distance >]	[{FUNC}]	[{< jeu, compétition, élections, etc. >}]	[NOM]	[ANM]	<i>wygrywać</i>	o	[ACC]	[< indication de distance >]	[{FUNC}]	[{FUNC}]	[{< jeu, compétition, élections, etc. >}]
[ANM]	<i>gagner</i>	[DET]	[< avantage > : < profit non- financier religieux > < demandant des efforts pour l'obtenir >]	-	-	[NOM]	[ANM]	<i>uzyskać</i>	-	[ACC]	[< avantage > : < profit non- financier religieux > < demandant des efforts pour l'obtenir >]	-	-	-

[<état, émotion, qualité, rire > de X]	<i>gagner</i>	[DET]	[ANM]	-	-	[NOM]	[<état, émotion, qualité, rire > de X]	<i>udzielać się</i>	-	[DAT]	[ANM]	-	-	-
[<idées >]	<i>gagner</i>	[DET]	[ANM]	-	-	[NOM]	[<idées >]	<i>szerzyć się</i>	<i>wśród</i>	[GEN]	[ANM]	-	-	-
[<idées >]	<i>gagner</i>	<i>parmi</i>	[ANM]	-	-	[NOM]	[<idées >]	<i>szerzyć się</i>	<i>wśród</i>	[GEN]	[ANM]	-	-	-
		[DET]	[<espace >]	-	-	[NOM]	[<incen- die, maladie, troubles, chô- mage >]	<i>rozszerzać się</i>	<i>na</i>	[LOC]	[<espace >]	-	-	-
[<incen- die, maladie, troubles, chô- mage >]	<i>gagner</i>	-	-	-	-	[NOM]	[<incen- die, maladie, troubles, chô- mage >]	<i>rozszerzać się</i>	-	-	-	-	-	-
[ANM]	<i>gagner</i>	[DET]	[ANM] [<demandant des efforts pour l'obtenir >]	<i>par</i>	{Z <qualité de X>}	[NOM]	[ANM]	<i>pozyskiwać</i>	-	[ACC]	[ANM] [<demandant des efforts pour l'obtenir >]	<i>dzięki</i>	[DAT]	Z <qualité de X >
[<émo- tion, rire > de Y > <sentiment /sensation négatifs >: < peur, angoisse, ennui, décourage- ment, fatigue, froid, sommell >]	<i>gagner</i>	[DET]	[ANM]	-	-	[NOM]	[<émo- tion, rire > de Y > <sentiment /sensation négatifs >: < peur, angoisse, ennui, décourage- ment, fatigue, froid, sommell >]	<i>ogarniać</i>	-	[ACC]	[ANM]	-	-	-

NP1	Verbe	Prép.	NP2	Prép.	NP3	NP1	NP1	Verbe	Prép.	NP2	NP2	Prép.	NP3	NP3
CO			CO		CO	cas	CO			cas	CO		cas	CO
[CONC: <voyageur> <véhicule>]	<i>gagner</i>	[DET]	[<espace>]	-	-	[NOM]	[CONC: <voyageur> <véhicule>]	<i>dotrzeć</i>	[FUNC]	[FUNC]	[<espace>]	-	-	-
[ANM]	<i>gagner</i>	[DET]	[<non-avantage> : <mauvais> : <maladie>]	[FUNC]	[ACTIVITE: <ALL>]	[NOM]	[ANM]	<i>dostawać</i>	-	[ACC]	[<non-avantage> : <mauvais>]	-	-	-
[ANM]	<i>gagner</i>	[DET]	[<espace>]	<i>pour</i>	Z <qualité de X>	[NOM]	[ANM]	<i>dostać się</i>	[FUNC]	[FUNC]	[<espace>]	<i>dzięki</i>	[DAT]	Z <qualité de X>
[ANM]	<i>Etre/se sentir gagné</i>	<i>par</i>	[<état, émotion, qualité, rire>]	-	-	[NOM]	[ANM]	<i>czuć się opanowanym</i>	<i>przez</i>	[ACC]	[<état, émotion, qualité, rire>]	-	-	-

Y. Wilks, discutant la question d'assignement du sens d'une occurrence donnée dans le discours à l'un des sens présenté dans le dictionnaire et la traitant de *very difficult* étudie le cas de EN *stake* dans les phrases du type :

I have a stake in this country
My stake on the last race was a pound

et pose la question si les deux occurrences de *stake* y représentent le même sens ou pas. La réponse que Y. Wilks propose est que cela dépend du fait si le dictionnaire sémantique contient un sens général – *vague* – pour *stake* (*Stake as any kind of investment in any enterprise*) à côté du sens *Stake as a post*, et si c'est ainsi la réponse est oui et sinon, et qu'il y a trois sens – *post, share, amount risked* – présentés pour *stake*, la réponse est non, puisque, somme toute, « word sense disambiguation is relative to the dictionary of sense choices available and can have no absolute quality about it » (Y. Wilks, 1972:12).

La question est quand même toujours de savoir combien de sens pour p.ex. *stake* il faut avoir dans un dictionnaire p.ex. EN – FR ou EN – PL, pour assurer une bonne WSD en vue d'une bonne TAO. La réponse s'impose d'elle-même dans le cas de la traduction : autant de sens qu'on puisse bien traduire *stake* en FR ou en PL et, en effet, il serait difficile de traduire les phrases citées ci-dessus de la même manière :

Mam udział w tym kraju / ¹*mam stawkę w tym kraju*
²*Mój udział w ostatnim wyścigu wynosił funta* / *Moja stawka w ostatnim wyścigu wynosiła funt*
J'ai une participation dans ce pays / ³*j'ai un enjeu dans ce pays* [possible, mais ne traduisant pas bien la phrase EN]
⁴*Ma participation dans la dernière course était un pound* [possible, mais ne traduisant pas bien la phrase EN] / *Mon enjeu dans la dernière course était un pound.*

La chose est d'ailleurs plus complexe, parce que même le sens concret de *stake* a au moins deux correspondants en FR et PL : *poteau* / *bûcher* et *slup* / *stos* et il serait impossible de traduire *stake*

In Medieval Europe many women were accused of being witches and were burnt at the stake [Cambridge Advanced Learner's Dictionary – CALD]

comme *poteau* ou *slup*.

Cela veut dire donc que, pour la TAO en particulier, mais aussi pour la traduction humaine, le dictionnaire bilingue devrait contenir au moins ces 4 sens et la traduction – perspective d'une autre langue – impose ce type de

répartition sémantique et on l'a bien vu ci-dessus d'une manière détaillée sur l'exemple de *gagner*.

Quelles en sont – s'il y en a – les conséquences pour la confection d'un dictionnaire monolingue? Rien d'impératif, parce que tout dépendra du but du dictionnaire et du public visé: on peut le faire comme le fait p.ex. *Collins English Dictionary* (CED) ou CALD, avec des distinctions très détaillées, surtout CED, ou comme *Longman Dictionary of Contemporary English* (LDOCE) ou *Wordwise Longman Dictionary*, avec 4 sens pour le LDOCE. Rien d'impératif aussi, parce que l'élaboration d'un dictionnaire monolingue doit se faire, au moins au départ, en fonction des distinctions de sens trouvées éventuellement au sein de la langue et non pas uniquement en fonction de traductions différentes dans une autre langue, pour éviter ce qu'on appelle *the sense-defined-by-language2 problems*. Mais, certainement, il n'est pas non-instructif, au contraire, de les regarder aussi de ce point de vue, parce que, comme le disaient I. Dagan, A. Itai et U. Schwall (1991), *Two languages are more informative than one*.

Par contre, une question tout à fait pratique se pose de savoir comment qualifier les contextes pour déterminer les paramètres pertinents pour la distinction entre p.ex. le sens de *participation* / enjeu ou encore *poteau* / bûcher.

On peut considérer ces paramètres:

- soit comme ce qui permet au lexicographe – *lumper*, qui n'a qu'un sens très général dans le dictionnaire, à la A. Wierzbicka p.ex., de faire relever le sens concret du mot utilisé dans le contexte donné au locuteur;
- soit comme ce qui permet au lexicographe – *splitter*, de voir la nécessité de distinguer différents sens du mot et de les mettre dans le dictionnaire pour «faciliter» la compréhension du mot au locuteur;
- soit comme ce qui permet au lexicographe, qui a pour but la TAO p.ex., de distinguer différents sens en fonction de différents ensembles de paramètres et de différentes traductions en langue-cible, assurant ainsi une WSD adéquate.

C'est justement la position qu'on a adoptée. Comme on l'a vu, on a utilisé dans ce but l'approche orientée objets avec la construction des classes d'objets. Les classes d'objets, du point de vue présenté ici, ne sont qu'une variante conceptuelle des catégories du type de celles qu'on trouve dans p.ex. un thesaurus, à ceci près qu'elles sont, d'une part, construites à partir de données linguistiques et non pas «ontologiques» et, d'autre part, beaucoup plus spécifiques et détaillées que les autres.

Leur application est plus étendue encore que ce qu'on a pu présenter dans un texte aussi court que celui-ci. Les classes d'objets ainsi conçues peuvent aider naturellement à classer tout élément dans un texte, p.ex. les éléments du contexte dépassant l'entourage immédiat du mot à traduire; c'est ainsi

qu'on peut déterminer p.ex. l'étendue de 10–50 de mots à classer dans une classe d'objets à gauche et à droite du mot en question et d'augmenter ainsi les chances d'une bonne WSD (cf. à ce propos p.ex. les analyses très intéressantes du type de D. Yarovsky, 1992, 1993; W. Gale, K. Church, D. Yarovsky, 1992).

Le débat donc du type *splitter* – plus de sens ou *lumper* – moins de sens n'est intéressant que pour l'élaboration des dictionnaires monolingues, et là encore est fortement relativisé au but du dictionnaire.

Pour la TAO la question n'est pas tellement pertinente et à la question posée combien de sens un mot de la langue-source a-t-il, on répondra, à la manière de ce sergent de la blague militaire expliquant les secrets du canon au simple soldat, autant qu'il faut – pour assurer la bonne WSD en langue-source, donc la bonne traduction en langue-cible, relativisant ainsi leur nombre au nombre de différentes traductions en langue cible.

Références

- Banyś W., 1997: «Approche classique et approche cognitive de la description lexicographique. Analyse d'une entrée de dictionnaire». *Neophilologica*, 13.
- Banyś W., 2002a: «Bases de données lexicales électroniques – une approche orientée objets. Partie I: Questions de modularité». *Neophilologica*, 15.
- Banyś W., 2002b: «Bases de données lexicales électroniques – une approche orientée objets. Partie II: Questions de description». *Neophilologica*, 15.
- Bar-Hillel Y., 1960: «Automatic Translation of Languages». In: F. Alt, A.D. Booth, R.E. Meagher, eds: *Advances in Computers*. New York: Academic Press.
- Dagan I., Itai A., Schwall U., 1991: «Two Languages are More Informative than One». *ACL*, 130–137.
- Dagan I., Itai A., 1994: «Word Sense Disambiguation Using a Second Language Monolingual Corpus». *Computational Linguistics*, 20.
- Desclés J.-P., Banyś W., 1997: «Dialogue à propos des invariants du langage (dans une perspective cognitive)». *Etudes Cognitives*, 2, 11–36.
- Fillmore Ch.J., Atkins B.T.S., 1992: «Towards a Frame-Based Lexicon: the Semantics of RISK and its Neighbours». In: A. Lehrer, E. Kittay, eds: *Frames, Fields and Contrasts*. New Jersey: Lawrence Erlbaum, 75–102.
- Gale W., Church K., Yarovsky D., 1992: «One Sense per Discourse». In: *Proceedings of the DARPA Speech and Natural Language Workshop*. New York: Harriman, 233–237.
- Gross G., 1996: *Les expressions figées en français: noms composés et autres locutions*. Paris: Ophrys, 161 p.
- Gross G., 1997: «Les classes d'objets et la désambiguïsation des synonymes». In: *Cahiers de Lexicologie*. Vol. 70. Paris: Didier Erudition, 27–40.
- Gross G., Clas A., 1997: «Synonymie, polysémie et classes d'objets». *Meta*, 42(1) [Presses de l'Université de Montréal], 147–155.

- Gross G., 1999a: «Elaboration d'un dictionnaire électronique». In: *Bulletin de la Société de Linguistique de Paris*. Vol. 94(1). Leuven: Peeters, 113–138.
- Gross G., 1999b: «La notion d'emploi dans le traitement automatique». In: *La pensée et la langue*. Kraków: Wydawnictwo Naukowe AP, 24–35.
- Hanks P., 1994: «Linguistic Norms and Pragmatic Exploitations or, why Lexicographers Need Prototype Theory, and vice versa». In: F. Kiefer, G. Kiss, J. Pajzs, eds: *Papers in Computational Lexicography: COMPLEX '94*. Budapest, 89–113.
- Idén, Véronis J., 1998: «Word Sense Disambiguation: The State of the Art». *Computational Linguistics*, 24(1), 1–40.
- Kilgarrriff A., 1993: «Dictionary Word Sense Distinctions: An Enquiry into their Nature». *Computers and the Humanities*, 26(1–2), 365–387.
- Kilgarrriff A., 1997: «I don't believe in word senses». *Computers and the Humanities*, 31(2), 91–113.
- McRoy S., 1992: «Using Multiple Knowledge Sources for Word Sense Disambiguation». *Computational Linguistics*, 18(1), 1–30.
- Ng H., Lee H., 1996: «Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach». In: *Proceedings of the 34th Meeting of the Association for Computational Linguistics (ACL-96)*. Santa Cruze: CA, 40–47.
- Wierzbicka A., 1989: *Semantics, Culture and Cognition*. Oxford: Oxford University Press.
- Wilks Y., 1972: *Grammar, Meaning and the Machine Analysis of Language*. London: Routledge.
- Wilks Y., 1978: «Making Preferences More Active». *Artificial Intelligence*, 11(3).
- Wilks Y., 1997: «Senses and Texts». *Computers and the Humanities*, 31, 77–90.
- Wilks Y.A., Sator B.M., Guthrie L.M., 1996: *Electric Words: Dictionaries, Computers and Meanings*. London: MIT Press.
- Wilks Y., Stevenson M., 1997a: «Combining Independent Knowledge Sources for Word Sense Disambiguation». In: *Proceedings of the Third Conference on Recent Advances in Natural Language Processing Conference (RANLP-97)*. Tzgov Chark, Bulgaria, 1–7.
- Wilks Y., Stevenson M., 1998a: «The Grammar of Sense: Using Part-of-Speech Tags as a First Step in Semantic Disambiguation». *Journal of Natural Language Engineering*, 4(2), 135–144.
- Wilks Y., Stevenson M., 1998b: «Word Sense Disambiguation Using Optimised Combinations of Knowledge Sources». In: *The 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL-98)*. Montréal, Canada, 1398–1402.
- Yarowsky D., 1992: «Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora». *COLING-92*.
- Yarowsky D., 1993: «One Sense per Collocation». *Proceedings ARPA Human Language Technology Workshop*, 266–271.